Posterior consistency in Bayesian inference with exponential priors

Masoumeh Dashti

University of Sussex

Seminar at Data Assimilation Centre, University of Potsdam

20 November 2020

Joint work with

S Agapiou (Cyprus), T Helin (LUT, Finland)

The setting

$Y^{(n)} \sim P_\theta^n, \quad \theta$ living in (Θ, d)

- prior μ on $\Theta \quad \longrightarrow \quad$ posterior $\mu(\cdot | Y^{(n)})$

Interested in asymptotic properties of $\mu(\cdot | Y^{(n)})$ as $n \to \infty$, assuming

- $\exists$ an underlying truth θ_0
- as $n \to \infty$, $Y^{(n)}$ corresponds to infinitely-informative data limit

---

Examples:

- White noise model: $Y_t^{(n)} = \int_0^t w_0(s)\, ds + \frac{1}{\sqrt{n}} B_t, \quad t \in [0, 1]$,
  $(\Theta = L^2[0, 1])$
- Density estimation: $Y_1, \ldots, Y_n \overset{\text{i.i.d}}{\sim} \pi_0, \quad \pi_0(x) = \dfrac{e^{w_0(x)}}{\int_0^1 e^{w_0(y)}\, dy}$

  $(\Theta = C[0, 1])$

Consistency with contraction rates

$\mu(\cdot|Y^{(n)})$ is said to contract with rate ϵ_n wrt d at θ_0 if

$$\mu\Big(\theta : d(\theta,\theta_0) \geq C\epsilon_n \Big| Y^{(n)}\Big) \to 0 \quad \text{in} \quad P^n_{\theta_0}\text{-probability}$$

Ghosal & van der Vaart 2007:

Conditions on prior

- μ puts sufficient mass around w_0,
- $\exists \Theta_n \subset \Theta$ with '*bounded complexity*' and $\mu(\Theta_n \setminus \Theta)$ sufficiently small

model: $\exists$ statistical tests distinguishing θ_0, θ_1 with error probabilities exponentially small in $d(\theta_0, \theta_1)$

satisfied by white noise and density estimation models

General contraction for white noise model

$Y_t^{(n)} = \int_0^t w_0(s)\, ds + \frac{1}{\sqrt{n}} B_t, \quad t \in [0,1]; \quad (\Theta, d) = (L^2[0,1], \|\cdot\|_2)$

Ghosal & van der Vaart 2007:

If there exist $\Theta_n \subset \Theta$ s.t. for ϵ_n with $n\epsilon_n^2 \geq 1$

- $\mu(\theta \in \Theta : \|\theta - w_0\|_2 \leq \epsilon_n) \geq e^{-cn\epsilon_n^2}$
- $\frac{\mu(\Theta \backslash \Theta_n)}{\mu(\theta \in \Theta : \|\theta - w_0\|_2 \leq \epsilon_n)} = o(e^{-n\epsilon_n^2})$.
- $\sup_{\epsilon > \epsilon_n} \log N(\epsilon/8, \Theta_n, \|\cdot\|_2) \leq n\epsilon_n^2$

Then posterior contracts at rate ϵ_n wrt $\|\cdot\|_2$ at w_0.

Outline

1. Posterior contraction with Gaussian priors
2. Posterior contraction for *p*-exponential priors
3. Contraction rates for white noise model

Outline

1. Posterior contraction with Gaussian priors
2. Posterior contraction for p-exponential priors
3. Contraction rates for white noise model

Contraction with Gaussian priors

van der Vaart & van Zanten (2008):

For appropriate ϵ_n

there exists $X_n \subset X$ s.t.

- $\mu(\|u - w_0\|_X < 2\epsilon_n) \geq \mathrm{e}^{-n\epsilon_n^2}$
- $\mu(X \setminus X_n) \leq \mathrm{e}^{-Cn\epsilon_n^2}$
- $\log N(\epsilon_n, X_n, \|\cdot\|_X) \leq Cn\epsilon_n^2$

Contraction with Gaussian priors

van der Vaart & van Zanten (2008):

For ϵ_n satisfying

$$\phi_{w_0}(\epsilon_n) \leq n\epsilon_n^2 \quad \text{with} \quad \phi_w(\epsilon) := \inf_{h \in H: \|h-w\|_X \leq \epsilon} \tfrac{1}{2}\|h\|_H^2 - \log \mu(\epsilon B_X)$$

there exists $X_n \subset X$ s.t.

- $\mu(\|u - w_0\|_X < 2\epsilon_n) \geq \mathrm{e}^{-n\epsilon_n^2}$
- $\mu(X \setminus X_n) \leq \mathrm{e}^{-Cn\epsilon_n^2}$
- $\log N(\epsilon_n, X_n, \|\cdot\|_X) \leq Cn\epsilon_n^2$

Given X and regularity of w_0, one can calculate ϵ_n
For appropriate models, rate ϵ_n is posterior contraction rate

Outline

1. Posterior contraction with Gaussian priors
2. Posterior contraction for p-exponential priors
3. Contraction rates for white noise model

p-exponential priors

- μ law of $(\gamma_\ell \xi_\ell)_\ell$ in $(\mathbb{R}^\infty, \mathcal{B}(\mathbb{R}^\infty))$, where

  $\gamma_\ell \to 0$ is a deterministic positive sequence

  $\xi_\ell \sim c_p \exp(-\frac{|x|^p}{p})$, i.i.d. $p \in [1, 2]$

  e.g. $(\gamma_\ell) \in \ell_2$ implies $\mu(\ell_2) = 1$

- For $X \subset \mathbb{R}^\infty$ a separable Banach space with Schauder basis $\{\psi_\ell\}$, μ can be identified with a measure on X through

$$u(x) = \sum_{\ell=1}^{\infty} \gamma_\ell \xi_\ell \psi_\ell(x)$$

  decaying properties of γ_ℓ
  and regularity of ψ_ℓ determine regularity of u

  e.g. $(\gamma_\ell) \in \ell_2$ and $\{\psi_\ell\}$ o.n. basis for $X = L^2$ implies $\mu(L^2) = 1$

- Space of admissible shifts of μ is

$$Q := \{h \in \mathbb{R}^\infty : \sum_{\ell=1}^{\infty} \frac{h_\ell^2}{\gamma_\ell^2} < \infty\}.$$

- For $h \in Q$

$$\frac{d\mu(\cdot - h)}{d\mu}(u) = \lim_{N\to\infty} \prod_{\ell=1}^{N} \frac{f_p(u_\ell - h_\ell)}{f_p(u_\ell)}, \quad \left(f_p(x) := c_p \mathrm{e}^{-|x|^p/p}\right)$$

$$= \lim_{N\to\infty} e^{\frac{1}{p}\sum_{\ell=1}^{N}\left(|\frac{u_\ell}{\gamma_\ell}|^p - |\frac{u_\ell - h_\ell}{\gamma_\ell}|^p\right)}.$$

Shepp (1965); Kakutani (1948); or see monograph *Bogachev (2010)*

- Let $\quad Z := \{h \in \mathbb{R}^\infty : \|h\|_Z := \sum_{\ell=1}^{\infty} |\frac{h_\ell}{\gamma_\ell}|^p < \infty\}$,
- $Z \subset Q$ for $p \in [1, 2]$, $\mu(Z) = \mu(Q) = 0$.
- For $p = 2$, $\quad Z = Q = H$
- When μ defined on function space X,
  $Z \subset Q$, both compactly embedded in X

- for $h \in Z$

$$\mu(\epsilon B_X + h) \geq \mathrm{e}^{-\frac{1}{p}\|h\|_Z^p} \mu_0(\epsilon B_X)$$

using $\frac{d\mu(\cdot - h)}{d\mu}(u) = \lim_{N\to\infty} \exp\left(\frac{1}{p}\sum_{\ell=1}^N \left|\frac{u_\ell}{\gamma_\ell}\right|^p - \left|\frac{h_\ell - u_\ell}{\gamma_\ell}\right|^p\right)$

Agapiou, D & Helin (2020)

Agapiou, D & Helin (2020):

For $\epsilon_n > 0$ with

$$\phi_w(\epsilon_n) = \inf_{h \in Z : \|h - w_0\|_X \leq \epsilon_n} \frac{1}{2} \|h\|_Z^p - \log \mu(\epsilon_n B_X) \;\leq\; n\epsilon_n^2$$

for any $C > 0$ there exists $X_n \subset X$ and $R > 0$ s.t.

- $\mu(\|u - w_0\|_X < 2\epsilon_n) \geq \mathrm{e}^{-n\epsilon_n^2}$
- $\mu(X \setminus X_n) \leq \mathrm{e}^{-Cn\epsilon_n^2}$
- $\log N(4\epsilon_n, X_n, \|\cdot\|_X) \leq RCn\,(\epsilon_n \vee \tilde{\epsilon}_n)^2$

Two-level Talagrand's inequality – 1994, $\forall M > 0$

$$\mu(A + M^{\frac{p}{2}} B_Q + M B_Z) \geq 1 - \frac{1}{\mu(A)} \exp(-cM^p)$$

$$X_n = \epsilon_n B_X + M_n^{\frac{p}{2}} B_Q + M_n B_Z, \quad M_n \asymp (n\epsilon_n^2)^{\frac{1}{p}}.$$

Approximating Q by Z: $\quad X_n \subset 2\epsilon_n B_X + \bar{M}_n B_Z$

Let $\quad h_1, \ldots, h_N \in \bar{M}_n B_Z \quad$ be $2\epsilon_n$-apart in $\|\cdot\|_X$

$$1 \geq \sum_{j=1}^{N} \mu(\epsilon_n B_X + h_j) \geq \sum_{j=1}^{N} e^{-\frac{\|h_j\|_Z^p}{p}} \mu(\epsilon_n B_X) \geq N e^{-\frac{\bar{M}_n^p}{p}} e^{-\phi_0(\epsilon_n)}.$$

Outline

1. Posterior contraction with Gaussian priors

2. Posterior contraction for p-exponential priors

3. Contraction rates for white noise model

White noise model

$$Y_t^{(n)} = \int_0^t w_0(s)\,ds + \frac{1}{\sqrt{n}} B_t, \quad t \in [0,1]$$

- $X = L^2[0,1]$
- μ : p-exponential, $p \in [1,2]$, $\gamma_\ell = \ell^{-\frac{1}{2}-\alpha}$, $\alpha > 0$ (α-regular)
- $w_0 \in B_q^\beta$, where

$$B_q^\beta := \{u \in \mathbb{R}^\infty : \sum_{\ell=1}^\infty \ell^{q\beta+\frac{q}{2}-1} u_\ell^q < \infty\}$$

Rates for white noise model

$$\phi_w(\epsilon_n) = \inf_{h \in Z : \|h - w_0\|_X \leq \epsilon_n} \frac{1}{2}\|h\|_Z^p - \log \mu_0(\epsilon_n B_X) \leq n\epsilon_n^2$$

with $Z = B^{\alpha + \frac{1}{p}}$.

- We find the fastest ϵ_n s.t. both

$$\inf_{h \in Z : \|h - w_0\|_X < \epsilon_n} \|h\|_Z^p \leq n\epsilon_n^2$$

and

$$-\log \mu(\epsilon_n B_X) \leq n\epsilon_n^2$$

Rates for white noise model

Agapiou, D & Helin (2020):

$w_0 \in B_q^\beta$; μ p-exponential and α-regular

- For $q \geq 2$

$$\epsilon_n = \begin{cases} n^{-\frac{\beta}{1+2\beta+p(\alpha-\beta)}}, & \text{if } \alpha \geq \beta, \\ n^{-\frac{\alpha}{1+2\alpha}}, & \text{if } \alpha < \beta. \end{cases}$$

- For $q < 2$ and $p \leq q$,

$$\epsilon_n = \begin{cases} n^{-\frac{2\beta q+q-2}{4(q-1)+4\beta q+2pq(\alpha-\beta)}}, & \text{if } \alpha \geq \frac{\beta p-1+a}{2p}, \\ n^{-\frac{\alpha}{1+2\alpha}}, & \text{if } \alpha < \frac{\beta p-1+a}{2p}. \end{cases}$$

- for $q \geq 2$, for $\beta = \alpha$ we get minimax rates $n^{-\frac{\beta}{1+2\beta}}$
- for $q < 2$, priors with $p < 2$ do better than Gaussian;

  Gaussian with linear estimators achieve $n^{-\frac{\beta-\gamma/2}{1+2\beta-\gamma}}$, $\gamma := (2-q)/q$.

- Rescaled priors for white noise model:

  $\bar{\mu}$: law of $(\lambda\gamma_\ell)_\ell$, $\quad \gamma_\ell = \ell^{-\frac{1}{2}-\alpha}$

  choosing λ_n appropriately gives

  minimax rate for $q = p < 2$ and $\alpha = \beta - \frac{1}{p}$

  minimax rate up to logarithmic factors
  for $p < q < 2$ and $\alpha = \beta - \frac{1}{p}$

- Density estimations

  Similar techniques give contraction rates,
  but they are sub-optimal